



Measuring Poverty Efficiently Using Adaptive Deprivation Scales

Nick Bailey¹

Accepted: 26 January 2020 / Published online: 31 January 2020
© The Author(s) 2020

Abstract

Deprivation scales are becoming increasingly familiar in research and official statistics on poverty. Taking advantage of the basis of these scales in Item Response Theory, this paper proposes a more efficient approach to implementation using adaptive testing. This maximises information collected for a given amount of survey time by screening respondents on the basis of initial answers, and halting questioning where it is very unlikely any further information will be gathered. The paper illustrates various implementations using data from eight years of the UK's *Family Resources Survey (FRS)*. Results show that an adaptive approach collects more than 99% of the information from the UK's official deprivation measure in half the survey time. In addition, the paper suggests improvements in the design of the UK's official deprivation scale as well as lessons for the development of future deprivation scales more generally.

Keywords Poverty · Deprivation · Item response theory · Adaptive measure · Household survey

1 Introduction

There has been a lengthy debate about how to best to measure poverty. Much of this focuses on money-metric approaches, and the relative merits of income- versus consumption-based measures (Gordon 2002; Meyer and Sullivan 2003; Brewer and O'Dea 2012). Income-based poverty measures suffer from a number of limitations, not least the volatility of current income compared with consumption, and the scale of both errors and biases in data collection (Moore et al. 2000; Meyer and Sullivan 2003; Brewer and O'Dea 2012). Consumption-based measures have been promoted in part because they may capture the position of poorer households better (Meyer and Sullivan 2003; Brewer and O'Dea 2012). Critics highlight that they are usually based on expenditures rather than consumption and may therefore miss many important aspects of living standards (Gordon 2002, 2006). There are also challenges with estimating current consumption of more durable items. A number

✉ Nick Bailey
nick.bailey@glasgow.ac.uk

¹ Urban Studies, School of Social & Political Sciences, University of Glasgow, 7 Lilybank Gdns, Glasgow G12 8RZ, UK

of non-money-metric approaches have been developed as alternative or complementary measures. These include subjective measures, asset-based measures, and deprivation scales (Gordon 2006; Brandolini et al. 2010). It is the last which are the focus of this paper.

A major disadvantage with money-metric measures from a practical viewpoint is that they are very time consuming and hence expensive to implement effectively (Meyer and Sullivan 2003; Deaton 2018). These measures are more accurate when information is collected across detailed sub-categories and when each household member is considered separately but that increases survey time. Furthermore, more reliable data are secured when respondents are asked to check documentary evidence or keep diaries rather than relying on recall but that again is more time-consuming. As a result, it can be difficult to justify the inclusion of detailed money-metric measures within general purpose survey where the primary focus is not material disadvantage. This potentially limits our knowledge of the incidence and impacts of poverty.

Deprivation scales offer one possible solution here. Unlike income measures which focus on resources, they capture actual living standards but unlike consumption measures, they focus on a small number of indicative items rather than a comprehensive picture. They have been developed through a body of research going back fifty years to Townsend's 1968/1969 study of poverty in the UK (Townsend 1979) and have now been implemented in numerous countries around the world, including high-, middle- and low-income contexts; see studies listed by Gordon (2006: p. 44), Dicks et al. (2009, p. 145) and Guio et al. (2017, p. 6).

Deprivation scales are also finding increasing application in official statistics. However, the versions used for these purposes tend to be greatly shortened compared to those produced in academic studies. The most recent academic study to update the UK's deprivation scale was the *Poverty and Social Exclusion UK (PSE-UK)* survey in 2012, producing a scale containing 44 items (Bramley and Bailey 2018). The UK's official measure of child poverty, however, includes a deprivation scale with just 21 items (DWP 2019a) while the EU's poverty measure for the *Europe 2020* targets includes a scale containing just 13 items (Guio et al. 2016).

The aim of this paper is to explore whether there is a more efficient way of implementing deprivation scales. Rather than simply shortening the scale and asking the same questions of everyone, it examines whether more information can be collected for a given amount of survey time using adaptive testing methods respondents may be asked vary subsets of the full scale. Adaptive testing is already widely employed in educational and psychological research (Embretson 1996; Embretson and Reise 2000). It has its basis in Item Response Theory (IRT) which argues that, with certain kinds of test, there is a strong pattern or ordering to responses to the items. As a result, information from responses to an initial subset of items can be used to determine which items should be asked next in order to maximise the information collected in a given time. With adaptive deprivation scales, the principle is the same but the approach is slightly different: initial responses are used to determine whether questioning should continue or halt if there is minimal prospect of gathering further information. It has long been recognised that deprivation scales are partly based in IRT but the logical step to adaptive testing does not appear to have been taken before.

There is an obvious concern that, in not asking all of the questions, adaptive scales may miss important information and lead to an underestimate of poverty levels. The paper therefore illustrates a range of approaches to implementing adaptive deprivation scales and assesses the trade-off between time saving and information loss. Information loss itself is assessed in a variety of ways, reflecting different potential uses for these scales. The paper

does this using data from the continuous household survey used to make the UK's official child poverty measures, the *Family Resources Survey* (FRS). In the process, the paper suggests improvements in the design of the FRS scale as well as lessons for the development of future scales more generally.

The next section of the paper introduces deprivation scales. It shows how the application of IRT in measurement supports the use of adaptive testing, and how IRT has been applied to deprivation scales but without the link to adaptive testing being made. The section finishes by outlining how adaptive testing could work with deprivation scales and discusses various ways in which we should assess its advantages and disadvantages. The following section introduces the data and methods used in the empirical section of the paper. The findings section reports on the analysis, examining the time savings and information loss from various ways of implementing an adaptive deprivation scale. The paper concludes with a discussion of the issues around use of this approach in practice as well as wider issues raised about the design of the current FRS deprivation scale and the criteria used to design all such scales.

2 Background

2.1 Deprivation Scales

Deprivation scales emerged from a succession of poverty studies in the UK: from Townsend's pioneering study (1979), through Mack and Lansley's (1985) incorporation of democratic or consensual approaches to further methodological refinements by Gordon and colleagues (Gordon and Pantazis 1997; Pantazis et al. 2006; Bramley and Bailey 2018). Mack (2018) provides an overview of the developments while Gordon (2018) and Guio et al. (2016, 2017) provide more technical details. In brief, respondents are asked about a series of items which are material goods or social activities. These items have been identified by a majority of the public as "necessities"—things which everyone should have and which no-one should have to go without because they cannot afford them. Respondents are asked to identify whether they have these goods or do these activities and, if not, whether that is because they cannot afford them or because they do not want them. For the UK measure, it is only a lack due to affordability which counts as deprivation. The number of items lacked is counted and a threshold set for identifying deprivation.

As measures of poverty, these scales may have conceptual as well as practical advantages. Conceptually, "[d]eprivation indices are broader measures [than those from consumption expenditure] because they reflect different aspects of living standards including personal, physical and mental conditions, local and environmental facilities, social activities and customs" (Gordon 2006, p. 38). In practical terms, they can be relatively quick to cover with questions which are easily understood by respondents without the need to refer to documents or records. In some applications, deprivation measures are used in conjunction with a low income test to give what Gordon (2002) terms a multidimensional poverty measure.

One criticism of deprivation scales comes from the adaptive preferences debate which questions whether those who have experienced poverty are able to give an 'objective' assessment of their current circumstances. In this case, the specific issue is whether poor individuals are more likely to report not wanting items in order to avoid having to state that they cannot afford them, creating a possible downward bias (Halleröd 2006). Empirical

evidence, however, suggests adaptation is not a significant issue. For example, Hick (2013) shows that allowing respondents to distinguish enforced lack from lack through preference leads to deprivation measures which are both more reliable and more valid, testing against a range of subjective and objective criteria. Crettaz and Suter (2013) find that some poverty measures with a subjective basis do appear to be affected by adaptation but that the kinds of deprivation scales examined here are largely unaffected. More generally, many deprivation studies have shown that the proportions of people giving “don’t want” as the reason for lacking tends to be very small (Gordon et al. 2000; Wright and Noble 2013; Bramley and Bailey 2018).

The apparent simplicity of deprivation scales belies their methodological sophistication. Discussing the construction of the current EU measure, Guio et al. (2016) describe four stages or tests which items have to pass before being accepted as part of the final scale: first, suitability or whether a majority of the population views each item as a ‘necessity of life’, and then tests of validity, additivity and reliability. The last is assessed using a combination of Classical Test Theory (CTT) and Item Response Theory (IRT). Under CTT, the overall reliability of the scale is assessed using Cronbach’s Alpha while the role of each item is assessed by omitting each in turn to see how Alpha changes. We discuss the use of IRT in relation to deprivation scales below, after introducing its role in testing more generally.

A separate issue concerns how to convert responses into an overall measure of deprivation. Three approaches can be identified. A simple count or score is used by the PSE-UK and the EU measures (Gordon 2018; Guio et al. 2016). Others have advocated using the value of the latent trait estimated by the IRT model but note that this appears to make very little difference in practice (Cappellari and Jenkins 2007; Fusco and Dickes 2008). Lastly, some have advocated prevalence weighting: counting the lack of an item which more people have or do as more important than the lack of an item which is less common. This is the approach adopted for the UK measure with the prevalence-weighted score re-scaled to run from 0 to 100 (see DWP 2018 for details). We return to the value of this below.

Despite the relative brevity of these scales, there are still challenges in implementing them as part of general purpose household surveys. The scales developed by studies such as the PSE-UK survey tend to be quite lengthy. In that case, the final version had 44 items for the full adult and child scale. For the purposes of official statistics where data are collected through household surveys which need to meet a variety of needs, these scales have been drastically shortened. The UK measure is produced from the *Family Resources Survey (FRS)* using a scale with just 21 items (McKay and Collard 2003) while the EU measure is produced through the EU-SILC datasets using just 13 items.

2.2 IRT and Adaptive Testing

IRT is based on the theory that there exists a latent (unobserved) factor which influences the responses an individual provides to each question or item in a test (Embretson 1996; Embretson and Reise 2000). From educational testing, an example would be how latent mathematical ability influences responses to questions in a maths exam. The responses each person gives enables their ability or level on the latent factor to be assessed but the full set of responses also enables the difficulty (also termed severity) of each item to be identified based on the proportion of people who get it correct (Edelen and Reeve 2007). Severity can therefore be used to select appropriate items for a given use. IRT also assesses the quality of items through their discrimination—the extent to which they are answered

consistently by people with similar abilities (Edelen and Reeve 2007). Low discrimination suggests there are factors other than the latent variable of interest shaping responses so the item is not well-suited to that test. IRT can be used to summarise the information provided by each item using an Item Characteristic Curve (ICC). This shows where on the scale of the latent variable it provides the most information. Items are most informative around the point where their severity lies. IRT also extends our understanding of the qualities of the test as a whole by revealing how the set of items produces more information at some points of the latent scale than others. In designing an effective test, knowledge about the intended uses of the scale can therefore inform the selection of items. In the case of a deprivation scale for the relatively affluent UK context, for example, we might assume that deprivation was experienced by a minority so the focus should be items where the severities lie in the more deprived half of the distribution and perhaps concentrate on the most deprived quartile. If our aim was to identify severe or extreme deprivation, we might want the information from the scale more focussed on the most deprived few percentiles.

IRT can also be used to make testing more efficient by tailoring or adapting tests while in progress. The set of items asked of each individual can be changed in real time, based on responses to previous questions. As Embreston (1996: p. 341) puts it, “better estimation of trait levels for all individuals are obtained from administering *different* test forms” (emphasis in original). No information is gained by giving students with low mathematical ability the most difficult questions, for example. The basis for most adaptive tests is a large set of items which have been pre-tested to establish their difficulty. Starting with items around the middle of the severity range of interest, questioning moves to more or less severe items depending on the responses. Each respondent is only asked a subset of full range of questions (those most informative about their ability) but all can be given an estimated ability on the same scale.

2.3 IRT and Deprivation Scales

The applicability of IRT to deprivation scales has been recognised for some time. A number of studies in English are cited here although Szeles and Fusco (2013) suggest the origins lie in studies published in French. One use of IRT is to determine whether items should be included in the scale or not, as noted above (Fusco and Dickes 2008; Guio et al. 2016; Gordon 2018). For Guio et al. (2016), for example, individual items are dropped if the severity is more than three standard deviations above the mean, indicating such extreme levels of deprivation that any household survey is unlikely to have a sufficient sample for analysis. They are also dropped if the discrimination is so low that it indicates a weak correlation with the latent trait; their threshold is 0.4. A second use is as a check that the scale is producing information at the appropriate level, by looking at the range of item severities (Szeles and Fusco 2013; Guio et al. 2016). A third use is to provide a more theoretically-informed basis for arriving at an overall measure of deprivation, by estimating the value of the latent trait for each respondent rather than relying on a simple sum or count. As noted above, what this actually tends to show is that the simple count works just as well and is therefore probably to be preferred on grounds of simplicity and transparency.

Despite these existing applications, there are three ways in which the use of IRT could be extended. First, while it is commonly used to check that items have severities in an appropriate range, it does not seem to be used to assess the information provided by the scale as a whole, as shown by Test Information Curves, for example. Reported item severities for deprivation scales often appear quite high, frequently concentrated between two

and three standard deviations above the mean (see Szeles and Fusco 2013 or Guio et al. 2016, for example). This indicates a concentration of information in the most deprived few percent of the range yet, for many analyses, we are interested in a much broader group; see for example, the annual analyses of poverty in the UK (Department for Work and Pensions (DWP) 2019a). This suggests a basic level of inefficiency due to excessive information being produced for one small part of the scale.

Second, but related to the first, while IRT is used to filter out items, it appears to play little role in the identification of items for potential inclusion in the first place. Where there is any discussion about the initial selection of items, the emphasis is usually on coverage of different domains or areas of consumption, reflecting theories about social necessities (Dickes et al. (2009) for the EU measure, or Gordon (2006) for the earlier PSE measure). However, if people give up consumption in certain domains before others as their incomes decline, a domain-drive approach to item identification can lead to a poor spread of information. Since item severity is strongly related to prevalence, it should be possible to make some assessment of this for most candidate items in advance and hence reach a judgement about the likely balance of information across the scale at the design stage.

Third, and the main focus of this paper, IRT provides a basis for constructing adaptive scales which are potentially much more efficient. Given deprivation scales in higher-income countries are designed to provide information at one end of the latent scale, the approach would be a little different to standard adaptive measures. Rather than starting in the middle of the range and moving up or down the severity scale, an adaptive deprivation scale would start with the least severe (most commonly lacked) items and then continue to more severe items only where there was a reasonable expectation that the respondent might lack those items.

Such an approach raises three questions: how should we identify when to halt questioning? How great are the potential savings in survey time? And how much information might be lost in the process? A number of approaches to implementing adaptive deprivation scales are outlined below, and time savings and information losses estimated for each. While the measurement of time saving is relatively simple (the proportion of questions unasked), different criteria can be proposed for assessing information lost depending on the intended purpose of the measure. The paper identifies three different potential purposes and reports on appropriate metrics for each.

3 Data and Methods

Data have been taken from the last eight waves of the UK's *Family Resources Survey* (2010/2011 to 2017/2018). The FRS is a continuous, cross-sectional survey running since 1993. It is the basis of a number of official statistics and analyses, including the annual UK Government report on *Households Below Average Incomes* (HBAI) (DWP 2019a). It collects detailed information on household incomes and other characteristics, and is used to produce a range of low-income poverty measures. Response rates are based on the proportion of households fully-complying with the survey (i.e. interviews with every non-dependent adult). For the period analysed here, they range between 52 and 62%, in general falling slowly over time (see Table M.1 in the analytical report for each year).

The material deprivation questions were initially introduced into the FRS in 2004/2005, with a set of 21 items based on research for the DWP by McKay and Collard (2003). That work drew heavily on the PSE 1999 survey and aimed to provide a measure which would

capture “most of the same people [as the full PSE scale] using a relatively short series of questions” (p8). The set of items was updated in 2010/2011 following McKay’s (2011) study which combined qualitative focus groups with parents and a national survey of opinions on deprivation items and prevalence. This recommended dropping four of the original items and replacing them with four new ones (Table 1). These changes were based in part on changes in public views about the suitability of items but also a desire to ensure balanced coverage of different domains. There was an acknowledgement that “there is only limited utility from having an item that is very rarely lacked” (p13) but nevertheless two of the new items were lacked by less than 2% of the population (children having fresh fruit and veg at least once a day and children having a warm winter coat).

The set has remained consistent since that study, so this paper focuses on the years from 2010/2011 onwards. In the current set, there are nine items on the living standards of the household as a whole and twelve on children’s living standards. The justification is that

Table 1 Items in the UK Family Resources Survey child deprivation scale

Item	2004/2005 set	2010/2011 set
<i>Adult/household item</i>		
Keep home in decent state of decor	Yes	Yes
Replace worn out furniture	Yes	Yes
Replace/repair broken electrical goods	Yes	Yes
Money to spend on self each week	Yes	Yes
Two pairs of all-weather shoes	Yes	
Regular savings of £20 a month	Yes	Yes
Household contents insurance	Yes	Yes
Keep home adequately warm	Yes	Yes
Able to keep up with bills*		Yes
Hobby or leisure activity	Yes	
Holiday one week a year	Yes	Yes
Family round for meal once a month	Yes	
<i>Child item</i>		
Fresh fruit/vegetables once a day		Yes
Bedroom for every child 10+ of different sex*	Yes	Yes
A warm winter coat		Yes
Garden or outdoor space*	Yes	Yes
Leisure equipment e.g. sports, bicycle	Yes	Yes
Hobby or leisure activity	Yes	Yes
Celebrations on special occasions	Yes	Yes
Friends round once a fortnight	Yes	Yes
Holiday away from home once a year	Yes	Yes
Toddler/nursery group once a week	Yes	Yes
School trip once a term	Yes	Yes
Activities e.g. drama, football etc.		Yes
Swimming once a month	Yes	
Number	21	21

See McKay and Collard (2003) and McKay (2011)

*Items where response categories are limited to yes/no

adults routinely shelter children from the worst effects of poverty so the identification of children in poor households requires a mix of child and household indicators (McKay and Collard 2003; McKay 2011).

With most items, the respondent is asked if the household/child has an item or does an activity and, if not, whether any lack is due to affordability. In three cases (marked ‘*’ in Table 1), the question is only about lacking the item or not. Some age-specific child items may be recorded as ‘not applicable’ for a given child (e.g. attending nursery or playgroup) and, following the Department for Work and Pensions (DWP) (2018) approach, this paper treats these as ‘not lacking’. The analysis is based on records for children to reflect the child population. Analyses use unweighted data unless otherwise noted.

Latent trait models (LTM) are the basis for this analysis. Following the PSE-UK study and Guio et al’s (2016) analysis for the EU, we use two-parameter models in order to capture variations in item severity and discrimination. LTMs model the logit (Y_{ij}) or the log of the odds that person i lacks item j as a function of individual position on the latent trait (θ_i), item discrimination (a_j) and item severity (b_j):

$$Y_{ij} = a_j\theta_i - b_j$$

Data were obtained from the UK Data Service (UKDS) as two linked data collections: the FRS data (DWP/NatCen/ONS 2019) and the associated HBAI dataset which contains derived variables used in the construction of the HBAI reports (DWP 2019b). The data are linked using unique reference numbers for each case. Dataset preparation was done using SPSS. Analytical work was done using R v3.6.0 (R Core Team 2013) with packages ‘dplyr’ (Wickham et al. 2019), ‘tidyverse’ (Wickham 2017), and ‘ltm’ (Rizopoulos 2006) current at May 2019. All the code (SPSS and R) are freely available from a github repository (<https://github.com/nick-bailey/Adaptive-deprivation-measure>). Data were obtained under a licence which prevents onward sharing but others may register with the UKDS, download the data under the same licence free of charge, download the code from github and so reproduce the results.

4 Findings

4.1 Descriptive Statistics

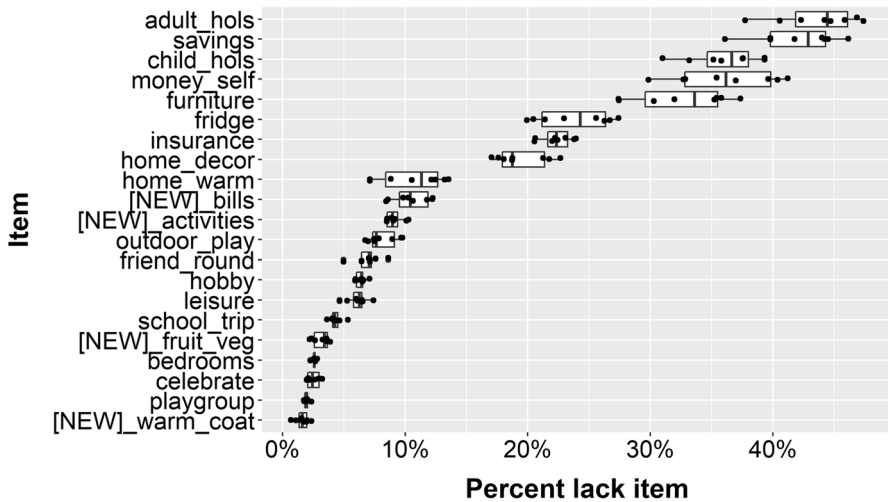
Around one in three children lacks none of the 21 items and more than half lack two items or fewer (Table 2). On the prevalence-weighted scale, a score of 25 or more (out of 100) is regarded as ‘deprived’. Just over one-in-five children meets this criteria. Children lacking seven items are almost always regarded as deprived. For later evaluation purposes, we also make a higher threshold for ‘very deprived’ on the DWP scale. Selecting the fairly arbitrary score of 40 or more, this covers around 1-in-12 children and corresponds to lacking 10 or more items in almost all cases.

The proportion of children lacking each item varies a great deal (Fig. 1). The proportions tend to fall over time in all cases so the order is very stable. The Figure shows which items households will typically give up as incomes fall. Holidays away from home for adults and children are among the first things to go, as are savings of £20 per month and, for adults, a small amount of money to spend on themselves each week. With the exception of child holidays, all of the first ten items are household- or

Table 2 Number of items lacked and deprivation rates—2010/2011 to 2017/2018

Number of items lacked	Percent of children (%)	Percent 'deprived' ($\geq 25/100$) (%)	Percent 'very deprived' ($\geq 40/100$) (%)
0	33.8	0.0	0.0
1	11.0	0.0	0.0
2	8.9	0.0	0.0
3	7.3	0.1	0.0
4	6.9	1.1	0.0
5	6.3	5.4	0.0
6	5.6	33.5	0.0
7	5.1	99.0	0.0
8	4.0	99.9	0.1
9	3.3	100.0	12.4
10	2.6	100.0	96.5
11	1.8	100.0	100.0
12+	3.4	100.0	100.0
All	100	22.4	8.1

Data for 87,842 children (unweighted)

**Fig. 1** Percent lacking each item each year—2010/2011 to 2017/2018. *Notes:* Points are percentages in each year lacking item while box plots summarise variation. The four items added in 2010/2011 indicated by “[NEW]”

adult-specific. Through falling savings, cuts to insurance and problems keeping up with bills, they indicate not just a deteriorating standard of living but also a rising financial vulnerability. Child-specific items are much less commonly lacked, reinforcing the findings from other research about how parents will try to shelter children from the worst

effects of poverty (Bramley and Bailey 2018). None of the child items is lacked by more than 10% of children.

4.2 Latent Trait Models

Latent trait models (LTMs) are run for each year of data. As noted above, two-parameter models provide item difficulty and discrimination scores for each year. Difficulty is a standardised (z) score which indicates the point in the distribution on the latent deprivation scale where we expect that half of our sample will lack a given item. This is more helpfully translated into percentiles (Fig. 2). As with the proportion lacking an item, difficulty scores are quite stable from year to year, and have been rising slightly over time. The ordering of the items on the basis of difficulty is therefore also very stable.

In general, there is a very strong relationship between difficulty (Fig. 2) and the proportion lacking an item (Fig. 1), the exception being ‘outdoor play’. A relatively high proportion lack this (around 5–9%, depending on the year) but the latent trait model gives it the highest difficulty rating. Further investigation shows that this item has the lowest discrimination in the set suggesting that there are other factors which influence responses, not just the latent factor that we are interested in (deprivation). This is the only child item with yes/no response categories (i.e. no additional test of lacking due to affordability) and this may lead to a number of higher income households having the item counted as a deprivation; we can speculate that these may be higher income households which have chosen for life-style reasons to live in relatively dense inner-urban locations where access to outdoor play space is restricted.

Figure 2 also shows that there are several items which are typically only lacked by children with quite extreme levels of deprivation—in the 98th or 99th centiles of the distribution. Given the focus in the DWP’s analysis on a threshold which corresponds to roughly the 78th centile (i.e. 22% are deprived on average—see Table 2), these items are likely

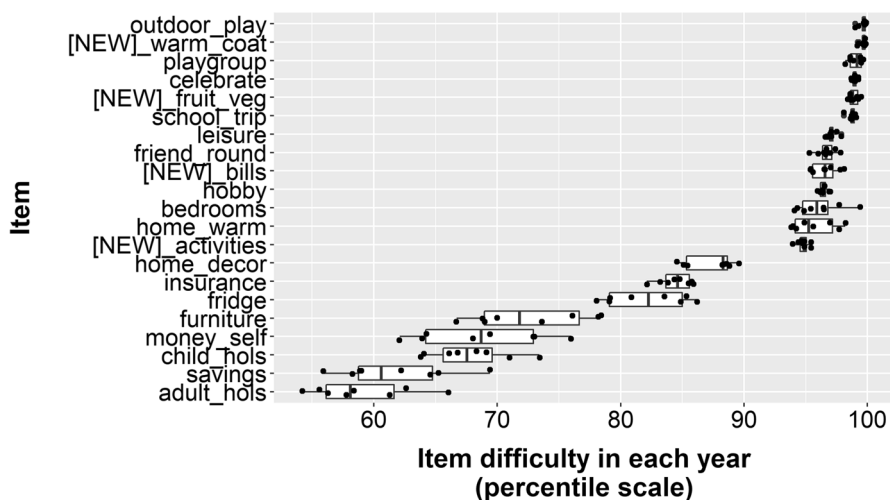


Fig. 2 Item difficulty by year—2010/2011 to 2017/2018. *Notes:* Points indicate difficulty rating in each year while box plots summarise variation

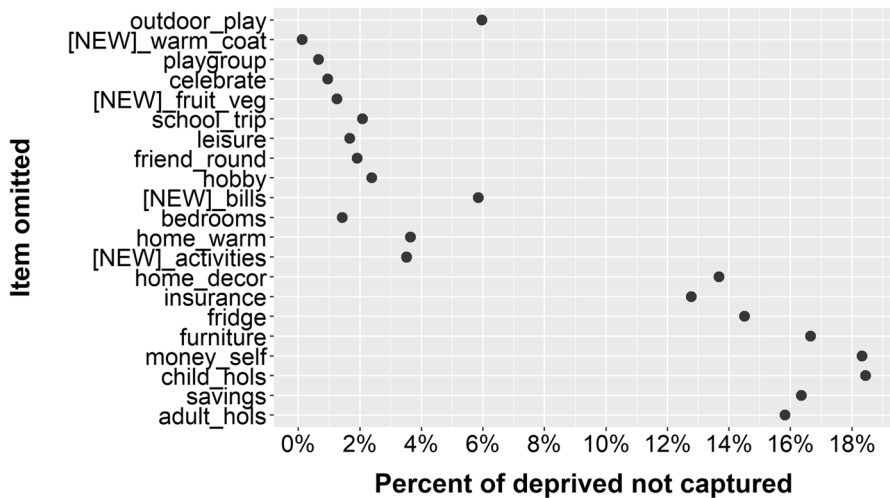


Fig. 3 Impact of dropping items from scale on proportion of deprived missed—2017/2018. *Notes:* Items ordered by difficulty as per Fig. 2

to play only a very limited role in distinguishing deprived from non-deprived households. Figure 3 demonstrates that this is the case. It shows the proportion of children regarded as deprived on the full scale who would no longer be regarded as such if we omitted each of the 21 items in turn from the prevalence-weighted index. The items with the lowest difficulty clearly play the strongest role in identifying households at the margin of being deprived as dropping any one of these would lead to the reclassification of 12–19% of the deprived. If we dropped the item ‘warm coat’, however, only 1-in-1000 of the cases identified as poor on the full scale in 2017/2018 would be missed. This item was one of the four added in 2010/2011. There are two more items (‘playgroup’ and ‘celebrate’) where their omission would mean a loss of less than 1% of the deprived and hence a change in the deprivation rate of less than 0.2 percentage points. For most items, the impact of dropping them is in proportion to their difficulty. Outdoor play is the most obvious exception and this is also the item noted above which has low discrimination.

The connection between IRT and deprivation scales can be used to show that prevalence weighting is unnecessary in theory. As items are strongly ordered, people will only tend to lack more prevalent items if they already lack less prevalent ones (i.e. in the language of IRT, they will only lack more difficult items if they already lack less difficult ones). People lacking more prevalent items do not need these to carry additional weight as a simple count would already give them a higher score. We see the effects of this in practice in the correlation between the simple count of items lacked and the prevalence weighted score. Across all eight years, it is 0.996 (range from 0.996 to 0.998, rising over the eight years). As others have already concluded, prevalence weighting is an unnecessary complication.

LTMs reveal where each item provides most information about the latent trait as described above. Item Information Curves give a visual summary of this (Szeles and Fusco 2013). Since item information is additive, the information yielded by the set of items as a whole can be summarised in a Test Information Curve. In Fig. 4, the left panel shows the usual representation of this, plotting information against standardised scores on the latent trait; this is a standard output from the ‘ltm’ package, for example. In this case, it shows

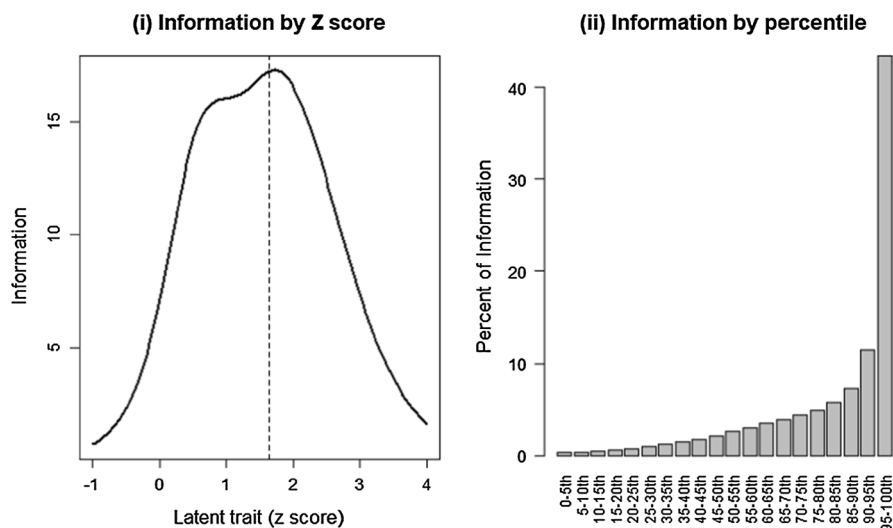


Fig. 4 Test information by (i) Z score and (ii) percentiles—2017/2018. *Notes:* Dashed line in left-hand pane shows 95th percentile of the distribution

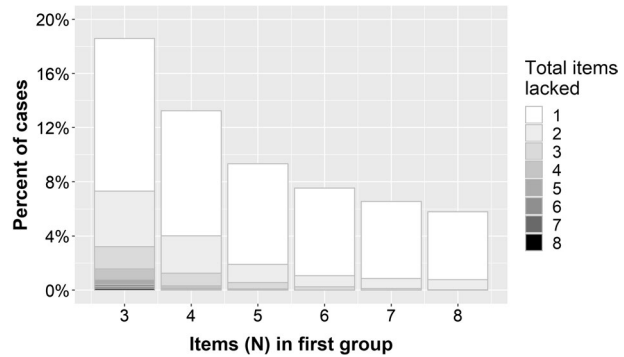
that the deprivation scale provides very little information around or below the average (zero on the latent scale) and this is quite appropriate: if we believe that fewer than half of all children in the UK are deprived, the scale should focus on the more deprived half of the distribution. The information provided rises sharply towards the higher values of the latent trait, peaking just below 2 standard deviations above the mean.

What is less obvious from this representation is the extent to which information is concentrated into the most extreme levels of the distribution. This is made a little clearer by the dashed line at the 95th centile of the distribution but is clearer still in the right panel which shows the same information curve transformed onto a percentile scale. In 2017/2018, almost half of all the information provided by this scale (46%) was in the most deprived 5% of the distribution. The proportion has been rising slowly over time, up from 42% in 2010. This does not appear well aligned to most users' interests which are in less extreme levels of deprivation.

4.3 Adaptive Scales

Unless we are particularly interested in severe deprivation, the skewing of the information curve leads to great inefficiency in measurement terms: most respondents answer negatively to all or almost all of the items as Table 2 above shows. However, this also means that an adaptive measure might capture the vast majority of the information provided by the full scale using a fraction of the survey time. For a simple version, we could ask about a group of items with lowest difficulty first (i.e. those most likely to be lacked) and halt questioning if answers suggested there was a very low probability that the person lacked any of the items with higher difficulty. Figure 5 illustrates this approach using an initial group of between three and eight questions, and assuming that we would halt questioning only if a household lacked none of the items in this initial group. Since

Fig. 5 Items lacked overall when lacking none of the first N items—2017/2018. *Notes:* Cases lacking no items are not shown



we would not have difficulty ratings available for the items in current survey year, we use the difficulties calculated from the previous year's data to give a realistic assessment of the approach here and throughout the rest of the paper.

For Fig. 5, we identify the children lacking none of the items in the initial group (those where we would halt questioning) and show the number of items lacked across the rest of the scale (the information we would miss). In the great majority of cases where none of the initial items is lacked, the child lacks none of the remaining items either (omitted from the Figure). The proportion is over 80% when we use an initial group of just three questions and over 95% when we ask an initial group of eight questions. For these cases, halting questioning does not lead to any loss of information.

For the remaining minority of cases, this approach does lead to some loss of information but the proportions are low and overwhelmingly composed of children who lack just one or two items. Losses decline as the number of questions in the initial group rises. For example, where households lack none of the first five items, fewer than one percent lacks three or more of the remaining items—and less than 0.1 percent lacks four or more. Given that households typically need to lack seven items to be regarded as deprived on the official measure, this approach would have almost zero impact on the estimated number of deprived children. The picture is very stable from year to year.

Table 3 Survey time savings and deprived cases missed—2017/2018

Items in initial group	Percent of cases lacking no items in initial group (%)	Survey time saving (%)	Percent 'deprived' missed ($\geq 25/100$) (%)	Percent 'very deprived' missed ($\geq 40/100$) (%)
3	50	43	0.4	0.1
4	47	38	0.1	0.1
5	45	34	0	0
6	44	32	0	0
7	44	29	0	0
8	43	27	0	0

'0%' means no cases. "Survey time saving" is the product of the proportion of questions not asked when questioning halts and the proportion of cases where questioning halts. "Percent of 'deprived' ['very deprived'] missed" is the proportion of cases identified as deprived [very deprived] by the full scale (lacking 7 [10] items or more) which would not be identified if questioning halted

The survey time saving with this approach can be measured by the proportion of cases where questioning would be halted early, multiplied by the proportion of items not asked. A large proportion of all cases would see questioning halt: from 43 to 50%, depending on the size of the initial set of items (Table 3), giving survey time savings between 27 and 43%. The saving falls as the size of the initial group rises, both because more questions are asked in the first group and because fewer cases meet the threshold of ‘lacking none’ for questioning to halt. Table 2 also provides two metrics for information loss: the proportion of people deprived or very deprived on the full scale who would not be identified as such if questioning halted. With an initial set of just three items, 0.4% of deprived cases would be missed by this approach. The missing cases quickly fall to zero with an initial set of five items and losses are less if we look at the higher ‘very deprived’ threshold.

It is possible to use a less cautious threshold to decide when to halt questioning, making greater time savings but losing more information. Figure 6 shows the trade-off for a range of scenarios, using the metric of proportion of deprived missed. Raising the threshold for stopping to one (i.e. halting where children lack one or none of the initial items), savings and information loss are both greater but information loss rises rapidly if the initial set is less than five items. With five items, the saving would be 43% but only 0.4% of deprived cases would be missed. With a threshold of two or less, information loss rises rapidly with fewer than seven items in the initial set and, overall, no better position is reached.

A more sophisticated adaptive measure divides the scale into successive groups of items, and makes the decision about whether to halt questioning after each group using progressively higher thresholds. At this point, the permutations multiply depending on the size of the groups and the various thresholds which might be applied at each stage. Indeed, permutations are practically endless since groups do not need to be of even size. Based on the results of the previous analysis, we focus on groups of five questions to illustrate this. Questioning might halt after the first group if a household lacks none of those items, or after the second group if they lack no more than one item, and so on. Figure 7 shows the savings and information loss from using thresholds of lacking zero, one or two items in the

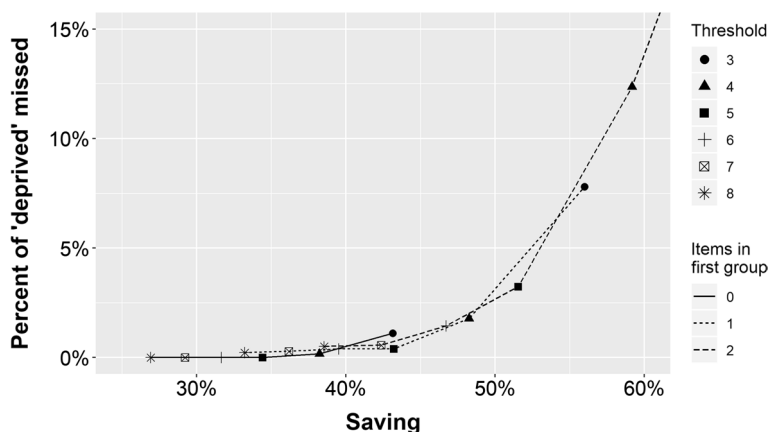


Fig. 6 Survey time saving versus information loss by size of first group and threshold—2017/2018. *Notes:* “Saving” is the product of the proportion of questions not asked when questioning halts at a given stage and the proportion of cases where questioning halts at that stage. “Percent of ‘deprived’ missed” is the proportion of cases identified as deprived by the full scale (lacking 7 items or more) which would not be identified if questioning halted at any stage

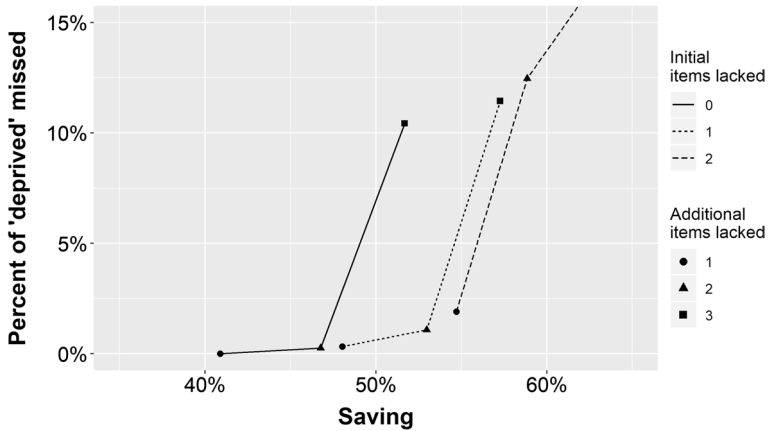


Fig. 7 Survey time saving versus information loss by adaptive measures—2017/2018. *Notes:* Questions asked in groups of five. “Saving” is the product of the proportion of questions not asked when questioning halts at a given stage and the proportion of cases where questioning halts at that stage. “Percent of ‘deprived’ missed” is the proportion of cases identified as deprived by the full scale (lacking 7 items or more) which would not be identified if questioning halted at any stage

first group (the three lines in the Figure). In each case, we show how these change when the threshold rises by one, two or three items with each group of questions.

A relatively conservative approach would be to use a threshold of lacking one or no items from the first five (the ‘initial items lacked’ threshold), and adding an additional one to the threshold at each subsequent stage. This gives a saving of just under half (48%) of the survey time required for this group of questions at a cost of missing just 0.3% of those who would be identified as deprived by the full scale—a modest but still worthwhile improvement on the simple approach above. An even more conservative approach would use an initial threshold of lacking zero items in the first set then one more with each subsequent set. This gives a saving of 41% with no deprived cases at all missed. Alternatively, an initial threshold of lacking two or fewer items in the first set and rising by one for each additional group gives a saving of 55% but misses 1.9% of deprived cases.

The saving/loss relationship is fairly stable from year to year and tends to improve over time. Using the first approach outlined in the previous paragraph (initial threshold of one, rising by one), savings range from 40 to 48% across the years, and the proportion of ‘deprived’ cases missed ranged from less than 0.1% to 0.9%. We can vary the size of the groups but there is no major improvement. For example, with groups of four and using the most conservative approach (an initial threshold of zero, rising by one), the saving would be 46% and loss 0.2%. With groups of six, an initial threshold of two and rising by one, the saving would be 49% and loss 1.0%. Achieving savings over 50% always leads to losses above 1.0%. In all cases, if the metric for information loss is the proportion of very deprived missed, losses are reduced.

4.4 Evaluation

Deprivation scales can be used in different ways and there are therefore different metrics to assess information loss. If the main interest is the analysis of deprivation as a binary category, we can measure information loss in terms of the proportion of the deprived missed by

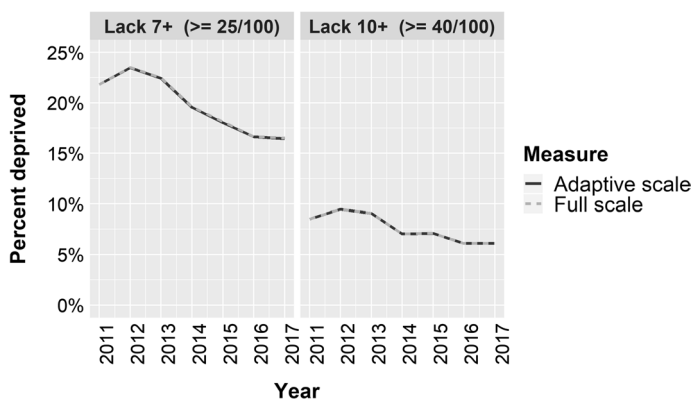


Fig. 8 Deprivation and severe deprivation rates using adaptive and full scales—2011/2012 to 2017/2018. Note: 2010/2011 cannot be shown since item severities from year before are not available

the adaptive scale as previously but this ignores the non-deprived who make up the majority and who are always correctly identified. A better alternative is to compare the estimated deprivation rate for the population using full and adaptive scales. Figure 8 illustrates this for two levels of deprivation, using the adaptive measure which produced time savings of 48% (based on groups of five questions, an initial threshold of lacking one or less, rising by one with each group). The estimated deprivation and severe deprivation rates from the adaptive scale are virtually indistinguishable from those produced by the full measure.

Alternatively, we may be interested in using the deprivation scale as a continuous measure so the relevant measure of quality would be the correlation between full and adaptive scales (Table 4). Again, using the same version of the adaptive scale, two correlations are shown: for all cases and for cases lacking at least one item on the full scale. In any year, between 30 and 41% of cases lack no items at all and there is a risk these might lead to an exaggeration of the level of correlation. In practice, the correlation is extremely high on both measures, and never less than 0.985.

Lastly, we may be interested in examining levels of deprivation item-by-item. This is the most demanding test and might be considered inappropriate: if we were really interested in

Table 4 Correlations between adaptive and full scale—2011/2012 to 2017/2018

Year	Correlation	Correlation (excluding zeroes)
2011	0.993	0.990
2012	0.993	0.990
2013	0.992	0.989
2014	0.993	0.989
2015	0.990	0.985
2016	0.991	0.987
2017	0.991	0.986

2010/2011 cannot be shown since item severities from year before are not available

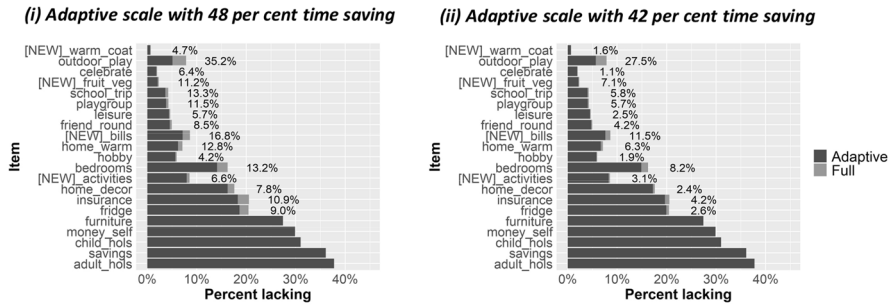


Fig. 9 Proportions lacking each item using adaptive and full scales—2017/2018

the proportions lacking specific items, we would probably not use the adaptive approach in the first place. Nevertheless, we show the proportions lacking each item using full and adaptive scales in Fig. 9. The top panel shows the proportions lacking each item using the same adaptive approach as before. Percentages show the proportion of people lacking each item who would be missed by the adaptive measure. For the first five items, the proportion is zero since everyone is asked these questions. Across the full set, the adaptive measure captures 95.2% of all items lacked. The worst performance is for ‘outdoor play’, an item already noted as problematic due to its low correlation with the latent trait. Ignoring that item, the proportion captured is 96.1%.

Of course, if the main interest were item-level analysis rather than the overall deprivation score, a more conservative approach could be taken in the set-up for the adaptive measure, saving less time but losing less information. As an illustration, the bottom panel of Fig. 9 shows the results for an approach which still yields a time saving of 42% (groups of five questions, stopping if lack none of first group, rising by one for each successive group). If we ignore ‘outdoor play’, this set up captures 98.3% of items lacked although there are still six items where at least 5% of lacking cases are missed.

5 Conclusions and Discussion

It has been recognised for some time that deprivation scales are based in part in IRT but the potential link to adaptive approaches does not appear to have been made before. This paper therefore provides a clear theoretical basis for the application of adaptive testing to deprivation scales and assesses a range of approaches to implementation in the context of the official UK deprivation measure. The paper identifies potential time savings and information losses for each, providing the basis for informed decision-making around adoption. The motivations for this work were in part improving survey efficiency and reducing respondent burden, but they were also about making it easier to justify the inclusion of deprivation scales in a wider range of household surveys so widening our knowledge about the incidence and impacts of poverty.

There is no single optimal design for an adaptive measure, rather there are value judgements about how much to prioritise time saving as against information loss. Furthermore, the design needs to be chosen with the intended uses in mind. Where the interest is in an overall measure of deprivation, the paper shows that adaptive scales can yield time savings approaching 50% yet still capture nearly all of the children identified as deprived on the

current full scale. Across the last seven years, the deprivation rate estimated on the basis of the adaptive scale was virtually identical to that from the full scale while the correlation between full and adaptive scales was around 0.99. Where the interest was in more detailed analysis of the proportions lacking each item, an approach which saves more than 40% of the time still captures 98% of all items lacked.

In the course of the analysis, the paper makes a number of observations about the existing UK measure. Looking at the individual items, one ('outdoor play') does not appear to function well, with poor discrimination and indications that non-material factors influence patterns of lacking. Other items in the scale provide almost no useful information for the majority of current analyses, notably 'warm coat' which was added in 2010/2011. There are strong grounds for removing these. More generally, updating of the items appears overdue given this was last conducted in 2010/2011. In terms of the calculation of deprivation scores, prevalence weighting appears quite unnecessary empirically and IRT shows why it is also inappropriate theoretically. The method represents a significant complication and hence cost in analytical time, as well as a barrier to wider understanding, and these seem strong grounds for discontinuing it.

Other recommendations concern the scale as a whole. The information produced by the current set of FRS items is heavily skewed to the most deprived few percent of the population. This reinforces the case for reviewing the current FRS scale, but it also highlights a weakness shared with many deprivation scales where insufficient attention is paid to item prevalence and hence severity at the time of selection. Greater attention needs to be paid to the intended uses when doing the initial identification of items for potential inclusion in scales in order to ensure the information produced by the final scale is better aligned to these uses. In most cases, this will mean making greater effort to identify items which are more widely lacked. A corollary is that less emphasis should be placed on covering different domains or areas of consumption when trying to identify potential items.

Acknowledgements This paper stemmed from work originally conducted as part of the *Poverty and Social Exclusion UK 2012* survey, funded by the Economic & Social Research Council (ES/G035784/1) and led by Professor David Gordon at the University of Bristol. The UK Government's Department for Work and Pensions has been responsible for the design and delivery of the FRS survey for the last 25 years. Access to data and documentation was provided by the UK Data Service. I am grateful to all of the above, although none bears any responsibility for the analysis or opinions presented here.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Bramley, G., & Bailey, N. (2018). *Poverty and social exclusion in the UK: Volume 2—The dimensions of disadvantage*. Bristol: Policy Press.
- Brandolini, A., Magri, S., & Smeeding, T. M. (2010). Asset-based measurement of poverty. *Journal of Policy Analysis and Management*, 29(2), 267–284.
- Brewer, M., & O'Dea, C. (2012). *Measuring living standards with income and consumption: evidence from the UK*. IFS Working paper W12/12. London: IFS.

- Cappellari, L., & Jenkins, S. P. (2007). Summarizing multiple deprivation indicators. In S. P. Jenkins & J. Micklewright (Eds.), *Inequality and poverty: Re-examined* (pp. 166–184). Oxford: OUP.
- Crettaz, E., & Suter, C. (2013). The impact of adaptive preferences on subjective indicators: An analysis of poverty indicators. *Social Indicators Research*, 114(1), 139–152.
- Deaton, A. (2018). *The analysis of household surveys: a microeconomic approach to development policy. Reissue edition with a new preface*. Washington: World Bank Group.
- Department for Work and Pensions (DWP). (2018). *Households below average income (HBAI) quality and methodology information report 2016/2017*. London: DWP.
- Department for Work and Pensions (DWP). (2019a). *Households below average incomes: An analysis of the UK income distribution 1994/1995–2017/2018*. London: DWP.
- Department for Work and Pensions (DWP) (2019b) *Households below average income, 1994/1995–2017/2018. [data collection]. 12th Edition. UK Data Service. SN: 5828. <http://doi.org/10.5255/UKDA-SN-5828-10>. UKDS.*
- Department for Work and Pensions, National Centre for Social Research, and Office for National Statistics: Social and Vital Statistics Division (DWP/NatCen/ONS) (2019) *Family Resources Survey, 2017–2018 [data collection]. UK Data Service. SN: 8460. UKDS.*
- Dickes, P., Fusco, A., & Marlier, E. (2009). Structure of national perceptions of social needs across EU countries. *Social Indicators Research*, 95(1), 143.
- Edelen, M. O., & Reeve, B. B. (2007). Applying item response theory (IRT) modeling to questionnaire development, evaluation and refinement. *Quality of Life Research*, 16, 5–18.
- Embretson, S. E. (1996). The new rules of measurement. *Psychological Assessment*, 8(4), 341–349.
- Embretson, S. E., & Reise, S. P. (2000). *Item response theory for psychologists*. New York: Lawrence Erlbaum Associates.
- Fusco, A., & Dickes, P. (2008). The rasch model and multidimensional poverty measurement. In N. Kakwani & J. Silber (Eds.), *Quantitative approaches to multidimensional poverty measurement* (pp. 49–62). London: Palgrave Macmillan UK.
- Gordon, D. (2002). The international measurement of poverty and anti-poverty policies. In P. Townsend & D. Gordon (Eds.), *World poverty: new policies to defeat an old enemy* (pp. 53–80). Bristol: Policy Press.
- Gordon, D. (2006). The concept and measurement of poverty. In C. Pantazis, D. Gordon, & R. Levitas (Eds.), *Poverty and social exclusion in Britain: the millennium survey* (pp. 29–70). Bristol: Policy Press.
- Gordon, D. (2018). Measuring poverty in the UK. In E. Dermott & G. Main (Eds.), *Poverty and social exclusion in the UK: Volume 1—The nature and extent of the problem* (pp. 17–40). Bristol: Policy Press.
- Gordon, D., & Pantazis, C. (1997). *Breadline Britain in the 1990s*. Aldershot: Ashgate.
- Gordon, D., Adelman, L., Ashworth, K., Bradshaw, J., Levitas, R., Middleton, S., et al. (2000). *Poverty and social exclusion in Britain*. York: JRF.
- Guio, A.-C., Gordon, D., Marlier, E., Najera, H., & Pomati, M. (2017). Towards an EU measure of child deprivation. *Child Indicators Research*, 11, 835–860.
- Guio, A.-C., Marlier, E., Gordon, D., Fahmy, E., Nandy, S., & Pomati, M. (2016). Improving the measurement of material deprivation at the European Union level. *Journal of European Social Policy*, 26(3), 219–333.
- Halleröd, B. (2006). Sour grapes: Relative deprivation, adaptive preferences and the measurement of poverty. *Journal of Social Policy*, 35(3), 371–390.
- Hick, R. (2013). Poverty, preference or pensioners? Measuring material deprivation in the UK. *Fiscal Studies*, 34(1), 31–54.
- Mack, J. (2018). Fifty years of poverty in the UK. In G. Bramley & N. Bailey (Eds.), *Poverty and social exclusion in the UK: Volume 2—The dimensions of disadvantage* (pp. 27–56). Bristol: Policy Press.
- Mack, J., & Lansley, S. (1985). *Poor Britain*. London: Allen and Unwin.
- McKay, S. (2011). *Review of the child material deprivation items in the Family Resources Survey. Research Report 746*. Sheffield: DWP.
- McKay, S., & Collard, S. (2003). *Developing deprivation questions for the Family Resources Survey. Working Paper 13*. Bristol: Personal Finance Research Centre, University of Bristol.
- Meyer, B. D., & Sullivan, J. X. (2003). Measuring the well-being of the poor using income and consumption. *Journal of Human Resources*, 38, 1180–1220.
- Moore, J. C., Stinson, L. L., Jr, W., & Edward, J. (2000). Income measurement error in surveys: A review. *Journal of Official Statistics*, 16(4), 331–361.
- Pantazis, C., Gordon, D., & Levitas, R. (2006). *Poverty and social exclusion in Britain: The millennium survey*. Bristol: Policy Press.

- Pantazis, C., Gordon, D., & Townsend, P. (2000) *Perceptions of the necessities of life: Scotland and England compared. Working Paper 24*. Bristol: Townsend Centre for International Poverty Research, University of Bristol.
- R Core Team. (2013). *R: A language and environment for statistical computing*. Vienna: R Foundation for Statistical Computing.
- Rizopoulos, D. (2006). ltm: an R package for latent variable modeling and Item Response Theory analyses. *Journal of Statistical Software*, 17(2), 1–25.
- Szeles, M. R., & Fusco, A. (2013). Item response theory and the measurement of deprivation: evidence from Luxembourg data. *Quality & Quantity*, 47(3), 1545–1560.
- Townsend, P. (1979). *Poverty in the UK: A survey of household resources and living standards*. Harmondsworth: Allen Lane.
- Wickham, H. (2017). *tidyverse: Easily Install and Load the 'Tidyverse'*. R package version 1.2.1. Retrieved January 13, 2020, from <https://CRAN.R-project.org/package=tidyverse>.
- Wickham, H., François, R., Henry, L., & Müller, K. (2019). *dplyr: A Grammar of Data Manipulation*. R package version 0.8.1. Retrieved January 13, 2020, from <https://CRAN.R-project.org/package=dplyr>.
- Wright, G., & Noble, M. (2013). Does widespread lack undermine the socially perceived necessities approach to defining poverty? Evidence from South Africa. *Journal of Social Policy*, 42(1), 147–165.

Publisher's Note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.